Northeastern

Department of Civil and Environmental Engineering
400 Snell Engineering
360 Huntington Ave.
Boston, MA 02115

CIVE 7381 Transportation Demand
Problem Set #3
Due: Wednesday, October 16, 2024

Problem 1
Consider a model that predicts the accident rates for different states in the U.S.:

$$Y_n = \alpha + \beta X_n + \varepsilon_n, \qquad n = 1,\ldots,50$$

where n are indices of states, X_n is the proportion of automobiles exceeding 55 miles per hour in state n, and Y_n is the number of fatalities per million vehicle miles.

The sample means of X and Y are 0.6 and 1.0 respectively. The sum of the squared deviations of X from its mean is 10.0, the sum of the squared deviations of Y from its mean is 1.0, and the sum of cross products of deviations of X and Y from their respective means is 2.0. In other words:

$$\sum_n (X_n - \overline{X})^2 = 10.0$$
$$\sum_n (Y_n - \overline{Y})^2 = 1.0$$
$$\sum_n (X_n - \overline{X})(Y_n - \overline{Y}) = 2.0$$

a) Compute the least squares estimates of α and β, the sum of squared errors (SSE), and R^2.

b) Test the hypothesis H_0: $\beta = 0$ against the H_1: $\beta \neq 0$.

Problem 2
Trip generation models for home-based social recreation purposes have been developed for a metropolitan area. Models have been developed separately for a household based on the number of workers in the household (WHH) and number of cars in the household (AO). The models for households with WHH = 1 and AO = 1 and households with WHH = 2+ and AO = 1 are given below along with the t-statistics (in parentheses) of the various parameters and their R^2 values.

	Model 1	**Model 2**
Parameter	**WHH = 1 and AO = 1**	**WHH = 2+ and AO = 1**
Constant (intercept)	1.014 (4.23)	
HHSIZE		-0.0625 (-1.99)
Log(HHSIZE)	0.3638 (0.98)	
Log(HHInc)		0.23 (6.78)
AreaType	-0.651 (-1.56)	
AreaType^2	0.252 (0.23)	
AreaType^3	-0.028 (-1.35)	
HBWTDUR		-0.00427 (-4.58)
Log(HBWTDUR)	-0.086 (-3.58)	
Sample Size	1720	500
R^2	0.0458	0.2021

Values in parentheses represent the t-statistics from the regression analysis.

HHSize = Household Size
LogHHSize = Natural Log of Household Size
LogHHInc = Natural Logarithm of Household Income in 1000s of $ (MAX(0, LN(Income))
AreaType = Density Based Area Type (see below for definition)
AreaType^2 = Density Based Area Type, Squared
AreaType^3 = Density Based Area Type, Cubed
HBWTDUR = Average One-Way Home-based work (HBW) trip duration, in minutes
LogHBWTDUR = Natural Log of Average One-Way HBW Trip Duration (max(0, *ln*(HBWTDUR))

To determine the area type, the area density is first calculated as:

*Area Density = (Total Population + 2.5 * Total Employment / Developed Acres)*

Then, based on the value of *Area Density,* the Area Type of a zone is determined according to the following table:

AREATYPE		*Area Density*
0	*Regional Core*	*> 300.0*
1	*CBD*	*100.0 - 300.0*
2	*Urban Business*	*55.0 - 100.0*
3	*Urban*	*30.0 - 55.0*
4	*Suburban*	*6.0 - 30.0*
5	*Rural*	*< 6.0*

The above models were calibrated using data from individual household surveys. Discuss and evaluate the two models in detail:

a) What are the explanatory variables and what do they intend to capture?
b) Is the impact of the explanatory variables reasonable (based on the estimation results)?
c) Comment on the t-statistics of the variables and their interpretation
d) How else you would include the variable *AREATYPE* in the model?
e) Comment on the R^2 values. Calculate the corrected R^2 for both models. How do they compare to the corresponding R^2?
f) What other variables would you include in the model (assuming the data is available)?
g) Does it make sense that different explanatory variables were used to capture the number of trips for the two different household types? Explain why yes, or why not.
h) Would you use these models for a planning study?

Problem 3

This problem is based on problem 2 from PS 2.
You are given a set of aggregate data from 57 traffic analysis zones (TAZ) in the Chicago Area (the data is quite old as you can infer from the "official" description of the variables included in the table in the next page). For each of the 57 zones you have available the average trips per occupied dwelling unit, the average car ownership, the average household size, and three zonal social indices. The data is the same data you used in PS 2.

Based on your preliminary analysis on PS 2, develop linear regression models that predict the auto trips/day generated by a household. Present the best two model specifications you came up with and interpret the parameters and their signs. Report all relevant results from statistical tests.

You may use any statistical software (including R and python) to estimate the models (you can also use Excel using the regression analysis function).

The data file contains the following variables that can be used for the specification of your model:

Name	**Description**
TODU	*Trips per Occupied Dwelling Unit* Trips refer to the daily frequency of person-trips via motor vehicle (auto driver or passenger) or public transit made from a dwelling unit by members of that dwelling unit. All trips whose origins were other than "from home" were ignored.
ACO	*Average Car Ownership* Cars per dwelling unit.
AHS	*Average Household Size* Number of residents per dwelling unit.
SRI	*Job/Skills Rank Index* This index reflects two elements: (i) the proportion of blue-collar workers, defined as the ratio of craftsmen, operatives, and laborers to all employees; and (ii) educational level as measured by the proportion of persons 25 years and older completing eight or fewer years of schooling. The index attains a maximum value when no residents are in the blue-collar jobs category, and all adult residents have more than eight years of education
UI	*Urbanization Index* This index reflects three elements: (i) the ratio of children under five years of age to the female population of childbearing age; (ii) percentage of women who are in the labor force, and (iii) the percentage of single units to total dwelling units. The degree of urbanization index would be increased by (a) lower ratio in (i), (b) higher percentage in ii); and (c) lower proportion of single dwelling units. High values for this index imply less attachment to the home because of fewer children, higher likelihood of women being employed, and less permanency of dwelling unit type in terms of average tenure.
MI	*Minority Index* This index is defined as the proportion of an area's residents who are minorities.